GHRC Data Management and Publication

Deborah Smith

Lucy Wang

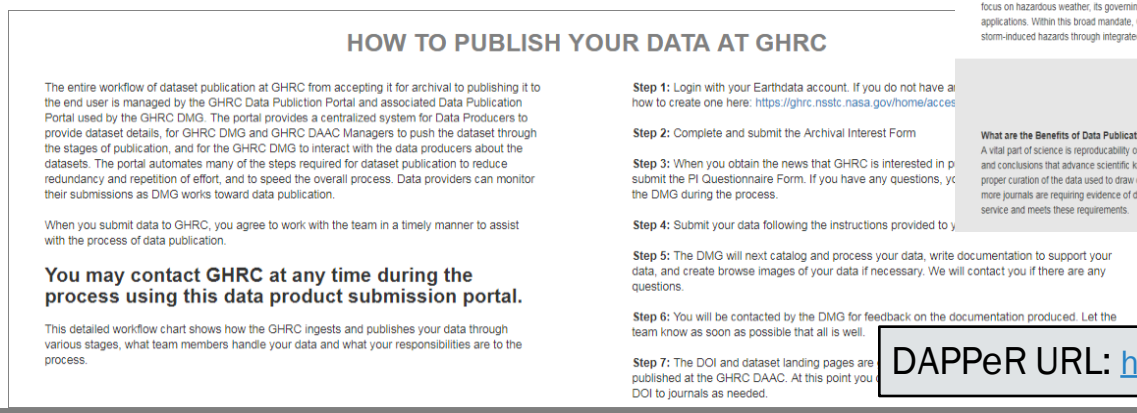
GHRC Data Management Team

2018 GHRC User Working Group Meeting
November 13-14, 2018

NASA
itsc
INFORMATION TECHNOLOGY AND SYSTEMS CENTER
THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

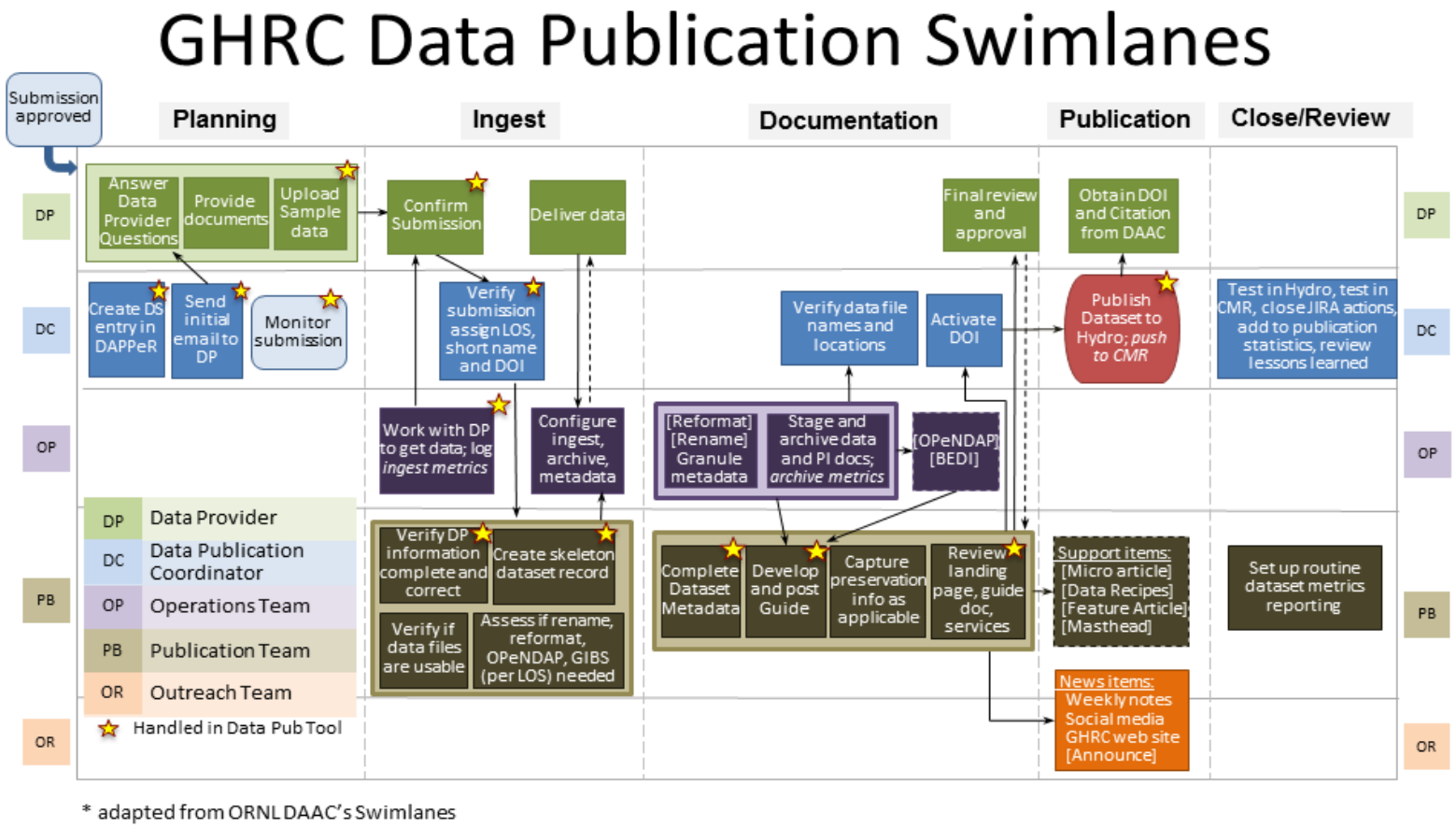
DAPPeR: GHRC Data Publication Tool

- DAPPeR is an online publication tool put into operation May 2017
- DAPPeR web page describes the publication process
- Feedback link provided so users can email GHRC with questions
- The secure portal interface has access for GHRC personnel and data providers



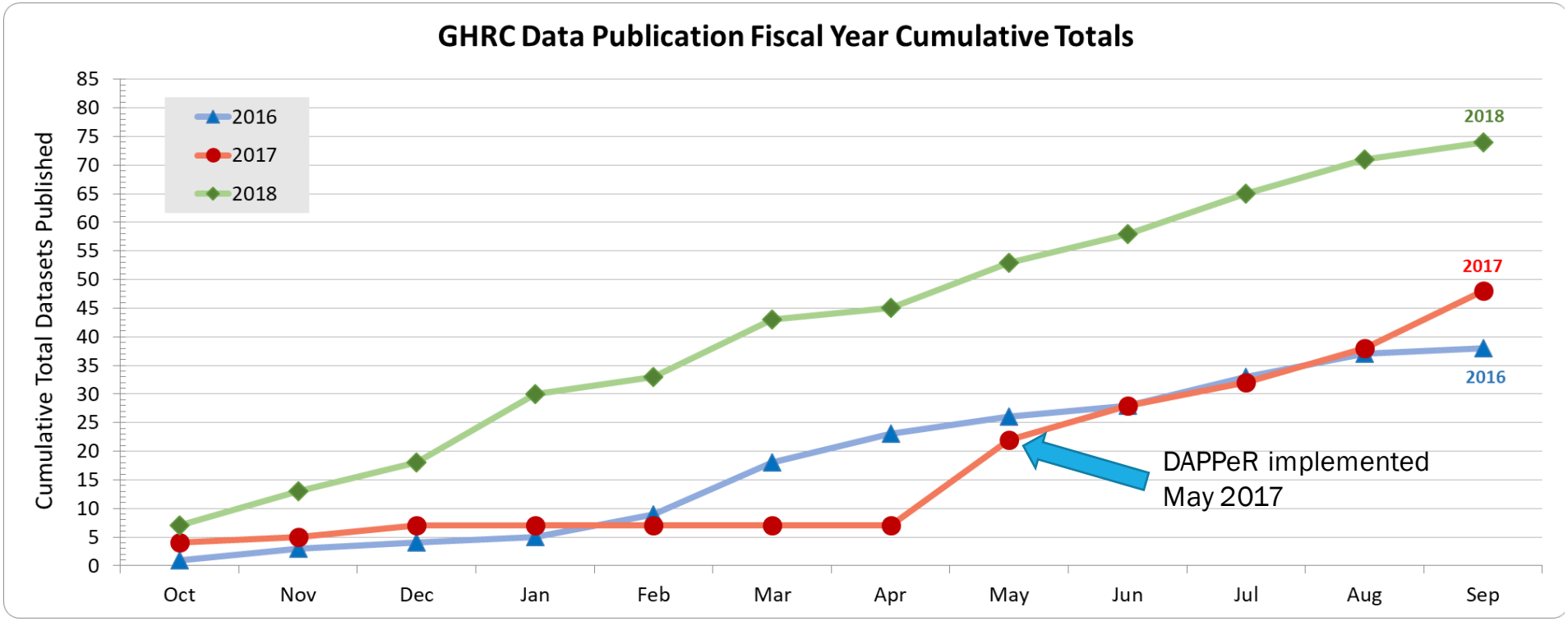
DAPPeR URL: https://ghrc.nsstc.nasa.gov/data-publication/

Data Publication Swimlane Chart Updated



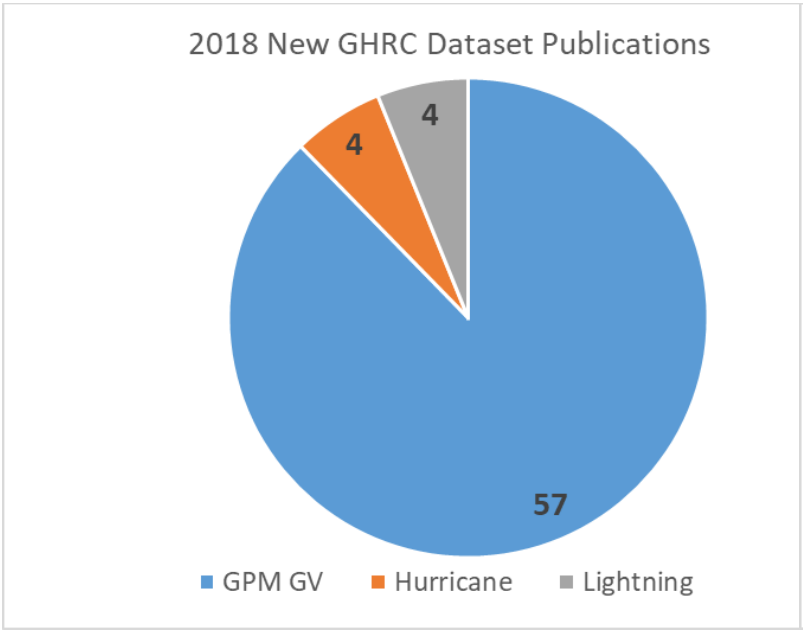
* adapted from ORNL DAAC's Swimlanes

FY2018 GHRC Data Publication Rate

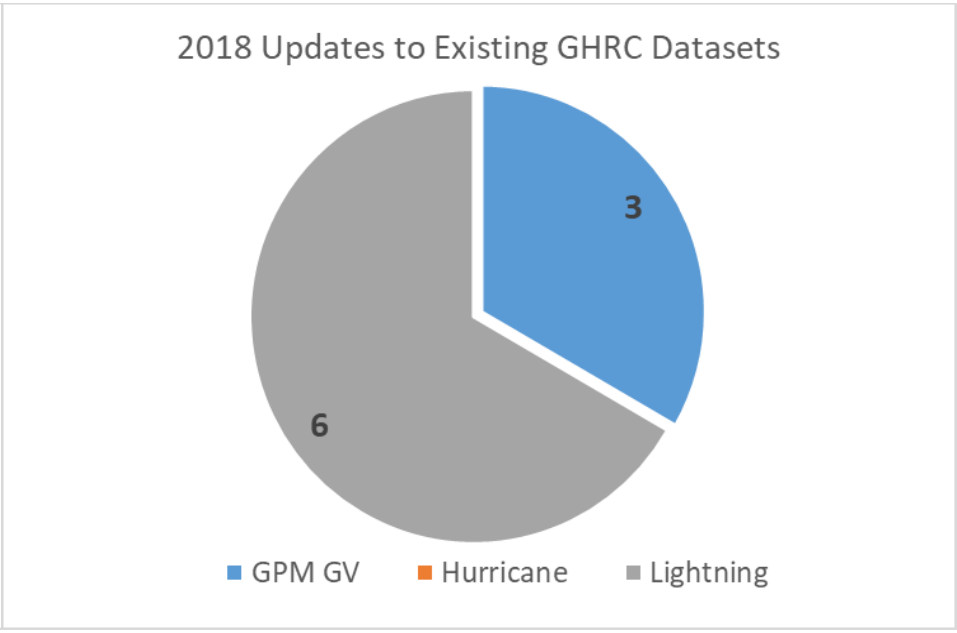


Dataset Types Published in FY2018

- 74 Datasets published in FY2018
- Averaged 6 datasets per month, double historical numbers



65 new datasets



9 dataset updates

GHRC Remaining Data to Publish

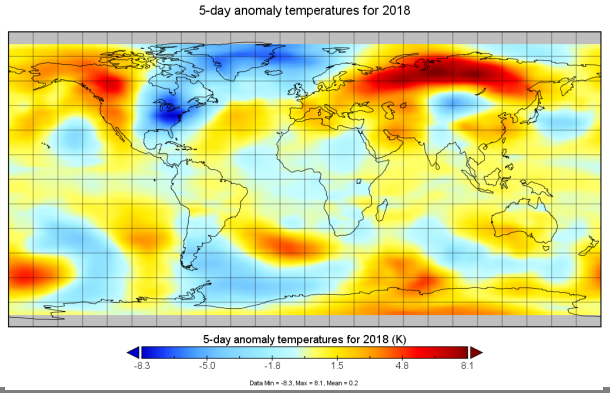
- Total of 19 unpublished datasets remain in queue
  - 9 GPM GV ancillary datasets - 96% of campaign complete
  - 3 HS3 support datasets - 81% of campaign complete
  - 7 lightning datasets - awaiting data for publication
    - validated ISS LIS datasets and lightning mapper array (LMA) data
- Plan to complete publication of remaining 19 datasets in first half of FY2019

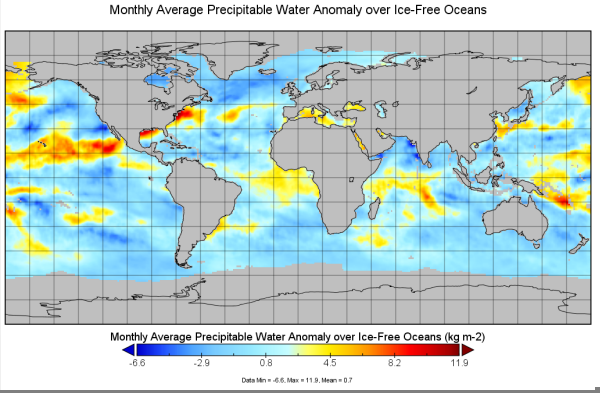
- **GHRC data holdings inventory in progress**

  Purpose is to:
  - Identify/confirm all data located on archives, public server, and in database
  - Ascertain our ability to rapidly replace public data from the archive in event of loss
  - Determine existing data organization, naming structure, storage format in order to assess problems and derive policies for better data handling in the future
  - Identify unpublished datasets delivered to GHRC drop servers, but not recorded

GHRC Ongoing Datasets

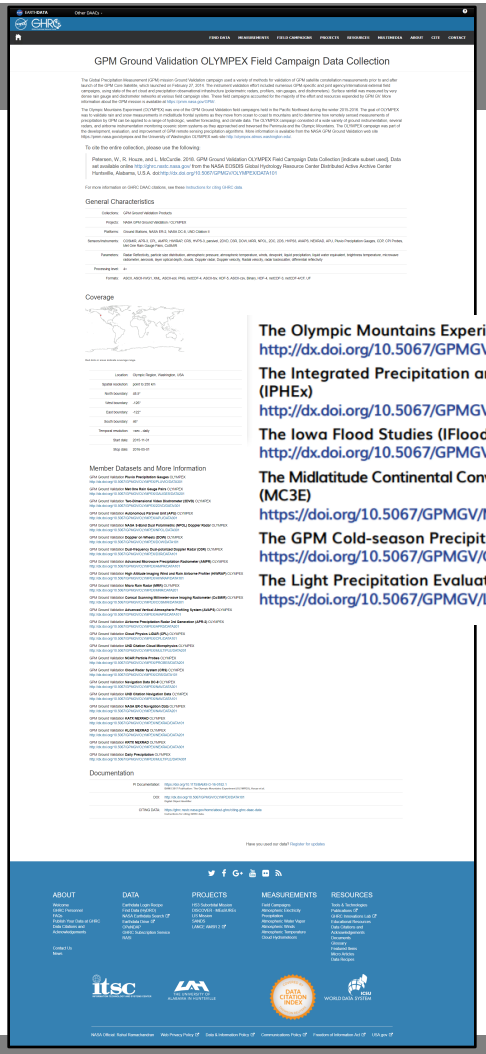
- Data that **continue to be collected** and made public on GHRC servers
  - ISS LIS data (2), NRT LIS data on LANCE
  - DISCOVER MEaSUREs SSMIS ocean products from F16 and F17 (8)
  - DISCOVER MEaSUREs TPW and Wind Speed climatology (2)
  - AMSU/MSU V6 Temperature Anomalies and Annual Cycle products (4)
- AMSU/MSU Atmospheric Temperatures updated to V6
  - New product: Temperature at the Tropopause (TTP)
  - Update already completed in October 2018 (FY2019)





New Field Campaign Collection DOIs

- Users frequently need to refer to the entire collection of data for a field campaign
- In FY2018, we added the DOIs for each GPM GV field campaign and the HS3 campaign
- "publication" of collections took effort
- Created a static landing page for each campaign with relevant information
- Obtained a campaign collection DOI that resolves to the static landing page
- Collection DOIs added to each collection dataset metadata and landing page
- Collection landing page contains a list of all member datasets for easy access



The Olympic Mountains Experiment (OLYMPEX):
http://dx.doi.org/10.5067/GPMGV/OLYMPEX/DATA101

The Integrated Precipitation and Hydrology Experiment (IPHEx)
http://dx.doi.org/10.5067/GPMGV/IPHEX/DATA101

The Iowa Flood Studies (IFloodS)
http://dx.doi.org/10.5067/GPMGV/IFLOODS/DATA101

The Midlatitude Continental Convective Clouds Experiment (MC3E)
https://doi.org/10.5067/GPMGV/MC3E/DATA101

The GPM Cold-season Precipitation Experiment (GCPEx)
https://doi.org/10.5067/GPMGV/GCPEX/DATA101

The Light Precipitation Evaluation Experiment (LPVEx)
https://doi.org/10.5067/GPMGV/LPVEX/DATA101

GHRC Metadata Repairs (ARC)

- ARC = Analysis and Review of CMR (Common Metadata Repository)
- Metadata = data about data, such as information that describes a dataset
- Metadata are needed for extracting and using data once published
- ARC project consists of a team of metadata checkers confirming quality of metadata for DAAC published data
- ARC report is sent to DAACs with problems/issues to be fixed and recommendations of changes to make
- GHRC received their ARC report in April 2016 and started collection-level metadata changes first
- Collection metadata changes are complete
- 63 dataset granule metadata changes remain (~30%)
- Remaining datasets are old, take more time, and need special handling
- Expect completion in FY 2019

User Guide and Dataset Documentation Improvements

- Old guides in html format
  - Not easy for data users to download and store with data
  - Not good for printing

- Now use Google Docs to construct, deliver as PDF
  - PDFs stored on server with data
  - Consistently formatted on any computer
  - However, not easy to mine for information

- In Process: conversion of old HTML guides to PDF versions. Time required to improve contents, increase accuracy, and add missing information or more recent information such as validation
- Lengthy process, of 459 datasets, ~25% completed

Contribution to Earthdata Pub Effort

- **Earthdata Pub**
  - A cross-DAAC effort to make the process of data publication at NASA DAACs more consistent for data providers, especially those interfacing with more than one DAAC
  - GHRC has been involved in effort throughout FY 2018 and will continue in FY2019

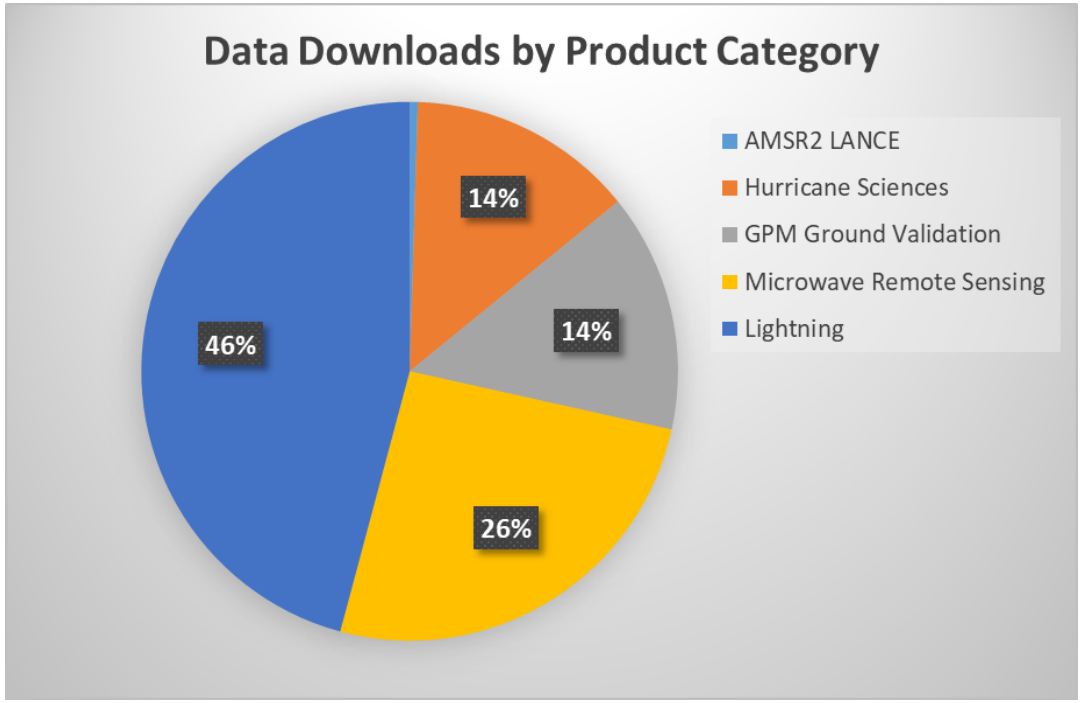
  - 2 subgroups:
    - Development of capability
    - Information and communication

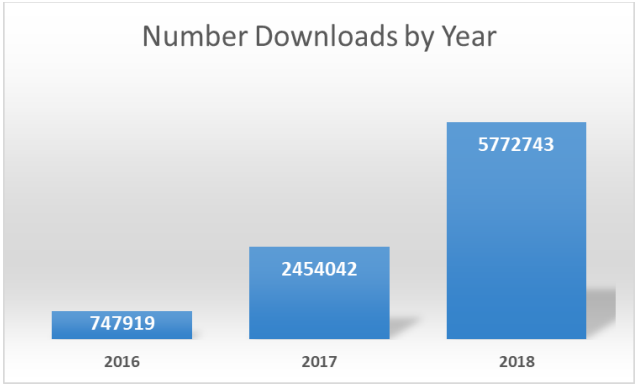
  - Information Subgroup
    - "Introduction to Data Publication at a NASA DAAC" guide in development
    - Synonym List to clarify terminology use across DAACs
    - Assessment of what information is asked of data providers by various DAACs to find similarities and to help with tool development

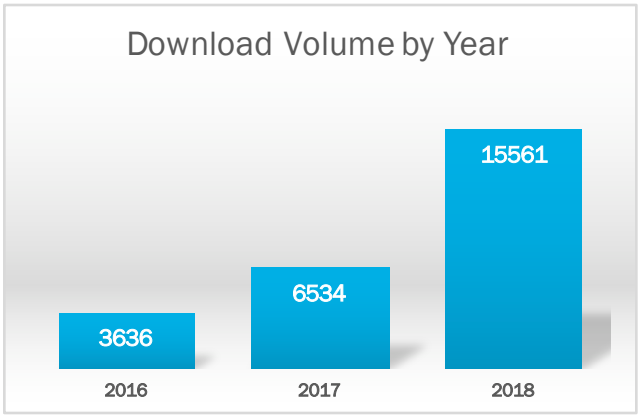
  - We anticipate GHRC to transition from DAPPeR to Earthdata Pub in the future

Data Access Metrics

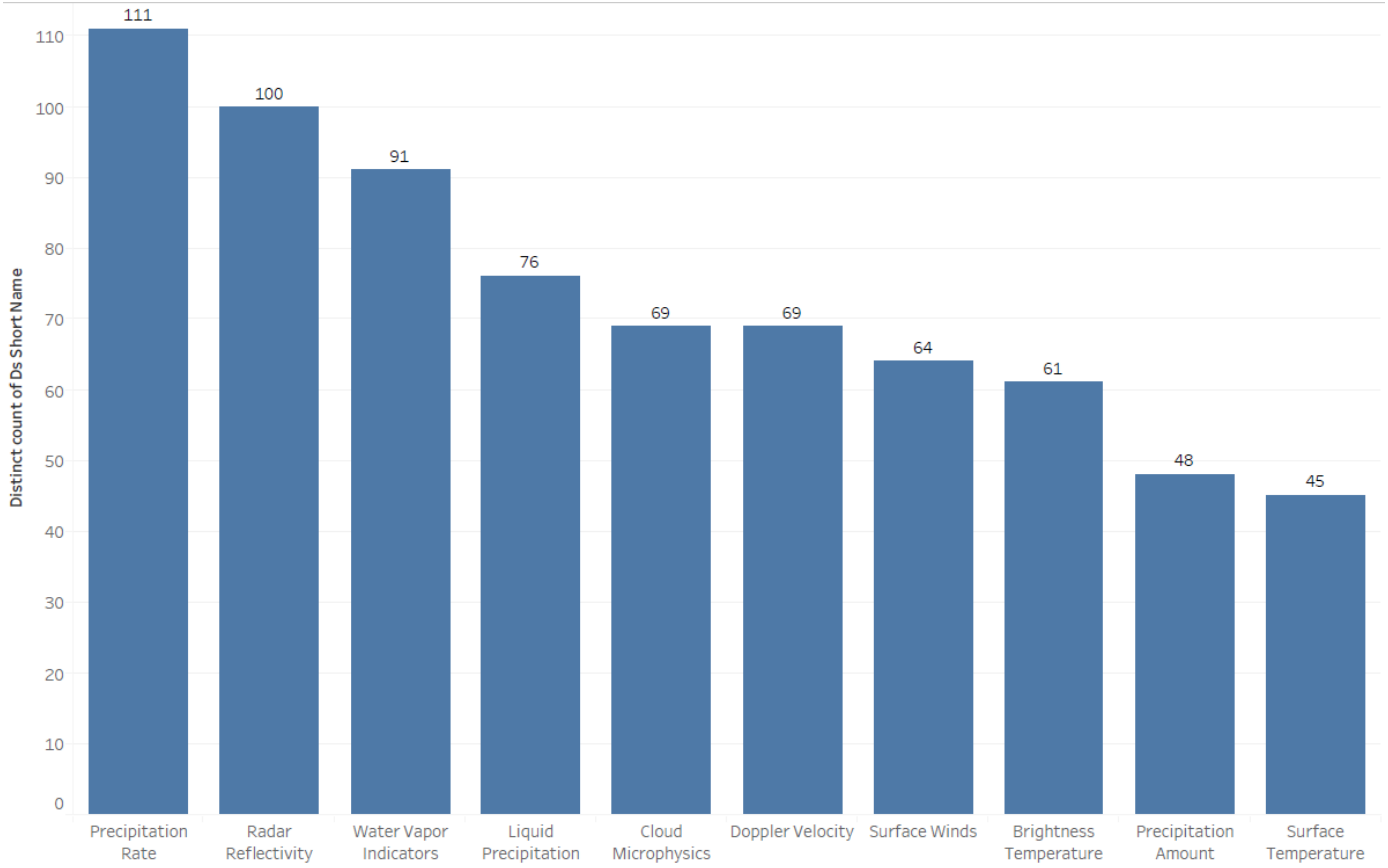
Total Number of Published Datasets at end of FY18: 444
Number Published Granules at end of FY18: >2.5 Million
Cumulative Downloads since Mar 2015: ~ 9 Million



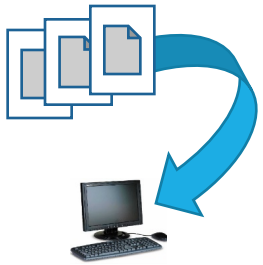




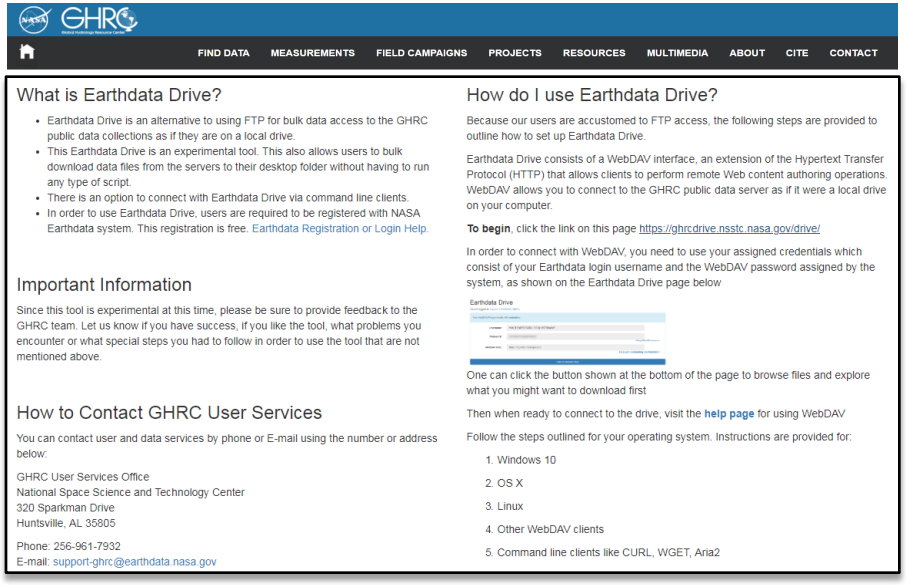
Data Holdings by Keyword



Bulk Data Download at GHRC



- Last year we reported on data access issues due to the loss of anonymous FTP
- At that time, we provided users with assistance for using WGET with a file of data URLs to download data via HTTPS. Some users have difficulty with Earthdata credentialing and WGET



GHRC

FIND DATA | MEASUREMENTS | FIELD CAMPAIGNS | PROJECTS | RESOURCES | MULTIMEDIA | ABOUT | CITE | CONTACT

What is Earthdata Drive?

- Earthdata Drive is an alternative to using FTP for bulk data access to the GHRC public data collections as if they are on a local drive.
- This Earthdata Drive is an experimental tool. This also allows users to bulk download data files from the servers to their desktop folder without having to run any type of script.
- There is an option to connect with Earthdata Drive via command line clients.
- In order to use Earthdata Drive, users are required to be registered with NASA Earthdata system. This registration is free. Earthdata Registration or Login Help.

Important Information

Since this tool is experimental at this time, please be sure to provide feedback to the GHRC team. Let us know if you have success, if you like the tool, what problems you encounter or what special steps you had to follow in order to use the tool that are not mentioned above.

How to Contact GHRC User Services

You can contact user and data services by phone or E-mail using the number or address below:

GHRC User Services Office
National Space Science and Technology Center
320 Sparkman Drive
Huntsville, AL 35805

Phone: 256-961-7932
E-mail: support-ghrc@earthdata.nasa.gov

How do I use Earthdata Drive?

Because our users are accustomed to FTP access, the following steps are provided to outline how to set up Earthdata Drive.

Earthdata Drive consists of a WebDAV interface, an extension of the Hypertext Transfer Protocol (HTTP) that allows clients to perform remote Web content authoring operations. WebDAV allows you to connect to the GHRC public data server as if it were a local drive on your computer.

To begin, click the link on this page https://ghrcdrive.nsstc.nasa.gov/drive/

In order to connect with WebDAV, you need to use your assigned credentials which consist of your Earthdata login username and the WebDAV password assigned by the system, as shown on the Earthdata Drive page below

One can click the button shown at the bottom of the page to browse files and explore what you might want to download first

Then when ready to connect to the drive, visit the **help page** for using WebDAV

Follow the steps outlined for your operating system. Instructions are provided for:

1. Windows 10
2. OS X
3. Linux
4. Other WebDAV clients
5. Command line clients like CURL, WGET, Aria2

- In April 2018, we implemented Earthdata Drive as our main alternative to FTP for bulk download
- Earthdata Drive consists of a WebDAV interface, a type of extenstion of the Hypertext Transfer Protocol (HTTP) that allows clients to perform remote web content authoring operations
- We provide instructions on the access page https://ghrcdrive.nsstc.nasa.gov/
- Limited metrics, some known users happy with functioning, no problems reported
- Other capabilities also developed (in later talk)

2017 ACSI Survey Feedback

ACSI Survey Feedback

- Respondents were **80% international**, 20% US
- Top user types of those responding were: **Earth Science Researchers (32%), University Grad/Undergraduate Students (29%), and Professors (16%)**

- **What events** were listed as important to users?
  - Hurricanes
  - Weather
  - Rain
  - Snowfall
  - Floods
  - Thunderstorms and severe weather
  - *Earthquakes*
  - *Solar storms*

In what **general areas/disciplines** did GHRC users say they used **Earth Science data?**

Landscape Architecture
Archaeology
Engineering
Disaster science
Environmental Studies
Forestry
Geography
Topography
River flow / Flood Study / River morphology
Remote Sensing

2017 ACSI Survey Feedback

ASCI User Response	Suggested GHRC Action
More GIS and image services	Increase GIS data recipes, more clearly point users to Worldview and BEDI, increase GIS knowledge at GHRC and provide intro guide to using GHRC data in GIS environments
Find data by country and improve spatial searching methods	Add region/country tags to datasets and add this search ability to HyDRO 2.0
Give a “related dataset” submenu on landing page	Define related dataset clearly for users. In addition to user guide related dataset section, consider adding to landing page
Better linking between documentation, tools, data	Improve web site documentation and resources pages (MUCH NEEDED!)
Improve search engine	Earthdata Search comparison with HyDRO 2.0
Provide data recipes for all datasets	Too much to do for all datasets, but could add a request form for new data recipes onto data the recipe web page asking users to submit wanted recipes
Provide webinars	Create a web page with descriptions of GHRC webinars and links, plus link to the NASA Webinar page for access to other webinars
Allow download of larger (longer in time) datasets without having to "click" each period/day	Better inform users of Earthdata Drive, make Earthdata Drive more visible, have a data download introduction document
Provide more related data, more world data	Expand GHRC data holdings, increase virtual collections with links to data at other DAACs
Highlight current images, water and weather related disasters	Post more ‘current events’ prominently on web page, tweet more and include tweets on web site

GHRC Data as Related to DAAC Mission Components

Mission Component	Collection	Program Type	Examples
Lightning	Lightning Products	Satellite sensors Ground measurements Airborne measurements	TRMM and ISS LIS OTD, LDAR, OLS, ACES, Field Mills
Precipitation Events	GPM Ground Validation Products	Airborne measurements Ground measurements	OLYMPEX, IFLOODS, IPHEx, GCPEX, MC3E, LPVEx
Hurricane/ Tropical Cyclone	Hurricane Sciences	Field campaigns	HS3, GRIP, CAMEX-3/4, NAMMA
		Satellite rain products	TRMM TCPF
Ocean Parameters, Atmospheric Temperatures	Microwave Remote Sensing Products	Satellite radiometer ocean products	SSM/I and SSMIS Ocean Products, RASI
		Satellite sounders	MSU, AMSU
Near Real-time Events	Operational/ Applications	LANCE	AMSR2 ISS LIS

FY2019 GHRC Data Publication Plans

Additional GPM GV data

1) ICE-POP

https://pmm.nasa.gov/ice-pop
https://wallops-prf.gsfc.nasa.gov/Field_Campaigns/ICE-POP/

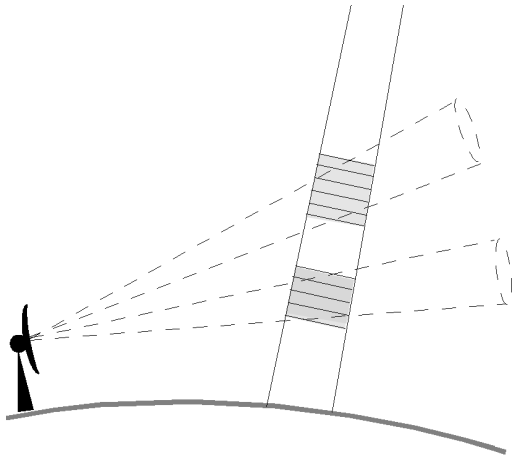
- International Collaborative Experiments for PyeongChang 2018 Olympic and Paralympic Winter Games
- Snow measurements from the start of the Olympics on Feb. 9 through the end of the Paralympics on March 18, 2018
- Focused on the measurement, physics, and improved prediction of heavy orographic snow in the PyeongChang region of South Korea
- Led by Korea Meteorological Administration
- International team (20 agencies from 11 countries)
- NASA data to be published:
  - Dual Frequency Dual Polarimetric Doppler Radar (D3R)
  - Micro Rain Radar (MRR)
  - Pluvio precipitation gauge
  - Parsivel disdrometer (APU)
  - Precipitation Imaging Package (PIP)
- Addition of NASA instrument data funded by GPM GV, Walt Petersen

FY2019 GHRC Data Publication Plans

2) GPM Validation Network (VN)

- Direct match-up of GPM's space-based DPR with ground radar (GR) data
- Vertical profiles with matching coincident DPR and GR data for precipitation events
- Key variables: radar reflectivity, precipitation rate, normalized intercept parameter
- Ongoing data since 2014
- Continental U.S., Alaska, Hawaii, Australia, Brazil.....
- GPM associated dataset important to GPM validation
- Addition of data funded by GPM GV, Walt Petersen

Schwaller et al., 2011 (https://doi.org/10.1175/2010JTECHA1403.1)

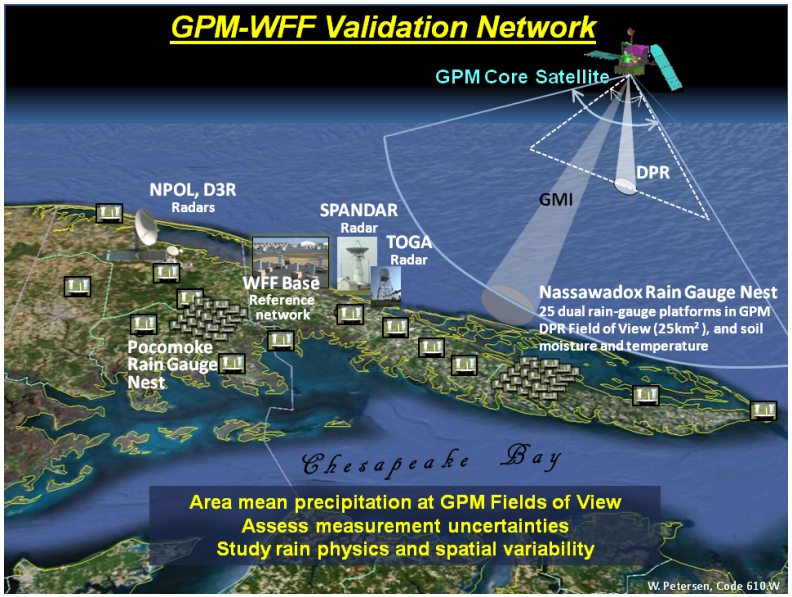


Example of DPR gate averaging at GR sweep intersections. Shaded areas show individual DPR gates intersecting the vertical extent of two GR sweeps (dashed) at different elevation angles. The reflectivity values of the individual DPR gates are averaged over the vertical extent of the GR sweeps, resulting in two matching volumes for the single DPR ray shown in this case.

FY2019 GHRC Data Publication Plans

3) WFF (Wallops Flight Facility)

- Instruments placed at Wallops or nearby area in between GPM GV field campaigns
- Ongoing data collection (need routine updates – 2x/year)
- Instruments include NPOL, D3R, MRR, rain gauges, APU
- Addition of this data is funded by GPM GV, Walt Petersen



FY2019 GHRC Data Publication Plans

Data reviewed by UWG in FY2018

1. Terra Fusion

   Was not advanced to ESDIS

2. GOES-R PLT Field Campaign Data
   - March to May, 2017
   - Advanced Baseline Image (ABI) and Geostationary Lightning Mapper (GLM)
   - Post-launch L1b and L2+ product validation

FY2019 GHRC Data Publication Plans

- Instruments onboard NASA ER-2 aircraft
  (1) Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)
  (2) Airborne Visible/Infrared Imaging Spectrometer Next Generation (AVIRIS-NG)
  (3) Cloud Physics Lidar (CPL)
  (4) Cloud Radar System (CRS)
  (5) Lightning Instrument Package (LIP)
  (6) Scanning High-resolution Interferometer Sounder (S-HIS)
  (7) Geostationary Coastal and Air Pollution Event (GEO-CAPE) Airborne Simulator (GCAS)
  (8) Fly's Eye GLM Simulator (FEGS)
  (9) ER2 navigation data
- Ground-based measurements
  (1) Red Lake, AZ: Mobile SURFRAD station
  (2) Ivanpah Playa, CA: Automated Solar Radiometer (ASR),
  ASD spectroradiometer and spectralon reference panel
  (3) Earth Networks Total Lightning Network (ENTLN)
  (4) Lightning Mapping Array (LMA):
  Colorado, Washington DC, NASA Kennedy Space Center, North Alabama, Oklahoma, West Texas, and Southern Ontario

What New Datasets for FY2020 and beyond?

- EVS-3 project selections have been announced

  Some of the selected projects may be appropriate to GHRC mission:
  - IMPACTS (Investigation of Microphysics and Precipitation for Atlantic Coast-Threatening Snowstorms) -*Intense snowfall events,* Lynn McCurdie of University of Washington
  - DCOTTS (Dynamics and Chemistry of the Summer Stratosphere) -*Impact of strong storms on stratosphere*, Ken Bowman of Texas A&M University

- SPoRT - Land Information System Land Surface and Soil Moisture Products

- Afternoon Brainstorming Session on future data recruitment

  We want your ideas on what GHRC can do to obtain more data relevant to our mission

Future of Data Publication at GHRC

- What will change as **GHRC heads into the cloud**?
  - Plans are for all of GHRC data holdings to be in cloud
  - Data publication process changes may be needed
  - Need to explore what data in cloud means with respect to GHRC activities and services
  - GHRC will participate this year in the new group working on Creating a Cloud Primer for New Users, spearheaded by Alaska Satellite Facility

  - **Afternoon brainstorming session to discuss your cloud ideas / issues /concerns that will help us prepare for the kind of guidance we need to offer data users.**

THANK YOU!

Questions?

2018 GHRC User Working Group Meeting
November 13-14, 2018

NASA

itsc
INFORMATION TECHNOLOGY AND SYSTEMS CENTER

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE